Computer Aided Audit Techniques and Tools (CAATTs)

Audit in general and fraud investigation in particular is getting complex day by day primarily due to migration to non-paper based systems. Processes and workflows are system driven and generally the entire control revolves around the computerised environment. Therefore, there is a clear trend that auditors would have to churn voluminous digitalised data as auditable records. Needless to say, auditors would therefore have to devise ways and means of verification that is commensurate with the changing times. The challenges are to examine various databases generated from different systems which may or may not be integrated.

Changing patterns of businesses, regulatory framework, scarcity of resources at auditors' disposal on one side and the ever increasing mountainous data on other hand is making audit a complex process. Use of CAATTs is, thus, indispensable to the Auditors and forensic accountants.

Let's see how we can use CAATTs in its simplest form for routine work.

Computer Aided Audit Tools or CAATs can be broadly divided into
- Generalised Audit Softwares (GAS) or
- Common Software Tools (CST)

Generalised Audit Software (GAS)
These are specialised softwares designed for accountants fostered with audit architecture in mind. The user interface is very simple for users to follow and with that objective, GASs very often have out-of-box-integration with leading accounting / other systems. However simple it may sound, it needs some training and experience to use them and some people do find it complex to operate. Another reason for GAS's unpopularity is due to its high cost. Some examples of GASs are ACL / Audit Command Language, Idea Data Analysis Software, Active Data, Wizrule, etc.

Common Software Tools (CST)
Due to shortcomings of GASs, CSTs have become popular over a period. Spreadsheets (like MS Excel, Lotus, etc.), RDBMS (like MS Access, etc.) and Report writers (like Crystal reports, etc.) are few examples of CSTs. Their widespread acceptability is due to its instant availability and lower costs. While spreadsheets may be extremely easy to use due to its simplicity and versatility, other CSTs may need some practice.

Whether one uses GAS or CST, it is imperative that the auditor is aware about the manner and processes that have lead to the data generation, the control environment revolving around the data and the source from where the data samples are imported into the GAS/CST.

As a part of computer-aided audit, an auditor needs to do one or more of the following.

A typical 8-point CAATT program :
1. **Check Missings**
2. **Check Duplicates**
3. **Round Numbers**
4. **Repetitive Odd-Numbers**
5. **Classification**
6. **Stratification**
7. **Single Transactions**
8. **Isolated Outliers**

1. Check Missing :

Here we basically try to identify the gaps in any workflow that has serial control mechanism. For example, missing cheques numbers, insurance policies numbers, bank fixed deposit receipts, good received notes, cash receipt numbers, etc. Depending upon the availability of data vis-a-vis audit objectives, an appropriate data-attribute (data field) may be selected to run this check. Missing gaps can be filtered for auditee's explanations.

> In one of the Company there were a few missing cheque numbers that could not be physically traced. Nor were proper explanations available whether they were factually cancelled or missing. Later it was revealed that the accountant had pilfered them and in fact one of the adjacent cheque number was surreptitiously encashed too. In another case, a missing Bank Guarantee document was misused to help an accomplice qualify for eligibility criteria in a Government Tender.

2. Check Duplicates :

This is converse of 'Check-Missings'--- the serial control numbers which ought not to be repeated are checked for duplicates. Thus all the documents (that have serial control), mentioned in the above paragraph can be checked for replication. Duplicated numbers could throw up serious gaps in the sourcing of these documents or in case they have been generated using some computerised system could mean a software bug. All such duplicated numbers need thorough detailed review to dispel any wrong doing.

Here auditor can also check duplication of clustered-fields. That is, say a cluster of vendor_code + bill_number can be checked for duplication, presence of which could indicate excess/double payments to same vendor. This is sometimes referred to as 'Same-Same-Different' check. In excel spreadsheet, the highlighting duplicate rule of the conditional format can easily spot the duplicate records.

3. Round Numbers :

Basically there is nothing wrong with Round numbers and it is not unusual to see many round number transactions in any commercial deals. However, sometimes round numbers are symptomatic of mysterious deals. Therefore the auditor should use some judgement to eliminate possible round number cases. For example, it is quite natural to generally spot round number transactions in monthly rentals, professional fees, audit remuneration, etc.- these transactions could be filtered out from the list of transactions with round numbers.

> Daily Times was a reputed newspaper publishing company that was publishing an eveninger-tabloid. Several round numbers payments were noticed to M/s. ABC Ltd., a news sprint supplier. ABC had been dealing with Daily Times for last 3 years and ever since payments to the vendors was always in round numbers. A detailed review and visit to ABC, revealed that there were several gaps in accounting of both the entities. Just to keep these gaps under the carpet, 'on account' payments in round numbers were affected to the vendor. Daily Times would adjust the bills per its own accounting system and ABC would follow theirs. Investigations exposed large unaccounted transactions of unsold newspapers returned back (for recycling).

4. Repetitive Odd-Numbers

This is converse of Round-numbers. Unlike the round numbers, repetition of odd numbers (particularly repetitions at decimals levels) are very rare coincident. Unless of course there is apparent reasons say, like for telco having promotional offer of Rs 199/- pre-paid packs – but in that case, the repetitions will be by volumes and not a few stray incidences here and there. Repeated odd-number transactions can be filtered for detailed verification and most often these will throw up some irregularities.

> In one MIS report of a public company running a restaurant business, the tobacco sale of Rs. 10,89,233/- was repeated consequently for the month of June and July. This was too much of coincident to catch the attention of the independent director who was otherwise a reputed investigator in his professional life. A few probing questions during the audit committee meeting revealed that the Manager was faking the MIS numbers and the amount was result of sloppy use of the spreadsheet's copy-paste command. The Manager had been manipulating MIS reports for quite some time and over the period casual complacent attitude had set-in. A detailed investigation launched subsequently exposed other mismanagement.

Repetitive checks can be performed on a given sub-set or across the entire set of population. For example, say in case of database of purchase records -- this check can be applied separately for each vendor or across all the vendors.

5. Classification

Classification is a process of arranging data into homogenous group or classes according to some common characteristics present in the data. This analysis aid the Auditor in getting a bird's eye view to see a panoramic whole of how the data is dispersed or where the concentration lies. Classification can be combined with other appropriate CAATT checks to enable more penetrative tests.

Classification could be done on various attributes (field) of data. For example, classify the data vendor-wise, salesman/agent wise, account wise, period wise, day-timing wise, etc. A typical classification of Purchase Data can be done to get following result

Sr	Party_Code	Count	Sum_Amount	%Count	%Amt
1	A-0001	3	7,375	0.09%	0.06%
2	A-0002	4	3,548	0.12%	0.03%
3	A-0003	107	46,430	3.26%	0.40%
4	A-0004	68	8,756	2.07%	0.07%
5	A-0005	1	188	0.03%	0.00%
			17,329	0.03%	0.15%
322	Y-0007	1			
323	Y-0008	1	447	0.03%	0.00%
324	Y-0009	3	26,989	0.09%	0.23%
325	Z-0001	21	5,940	0.64%	0.05%
326	Z-0002	13	9,496	0.40%	0.08%
327	Z-0003	2	165	0.06%	0.00%
328	Z-0004	1	56	0.03%	0.00%
Total Vendors =328		**3283**	**1,16,83,516**	**100.00%**	**100.00%**

The above list running into several pages can be further filtered for say %Count<1% and %Count >0.5% to get a very small sample size as follows :

Sr	Party_Code	Count	Sum_Amount	%Count	%Amt
16	A-0016	20	4,26,636	0.61%	3.65%
23	A-0023	13	15,94,532	0.40%	13.65%
34	A-0034	28	88,301	0.85%	0.76%
73	A-0073	3	99,358	0.09%	0.85%
96	A-0096	7	61,264	0.21%	0.52%
145	D-0002	14	1,66,954	0.43%	1.43%
149	D-0006	7	1,36,835	0.21%	1.17%
155	E-0006	7	2,30,400	0.21%	1.97%
179	I-0001	4	1,47,862	0.12%	1.27%
185	I-0007	10	68,497	0.30%	0.59%
207	L-0001	8	3,41,660	0.24%	2.92%
227	M-0019	4	5,09,856	0.12%	4.36%
238	N-0004	6	67,089	0.18%	0.57%
257	P-0009	1	1,07,361	0.03%	0.92%
260	P-0012	1	59,967	0.03%	0.51%
290	T-0004	12	3,02,458	0.37%	2.59%
301	T-0015	28	4,27,933	0.85%	3.66%
306	U-0003	1	1,52,343	0.03%	1.30%
Total Vendors = 18		**174**	**49,89,306**	**5.30%**	**42.70%**

As can be seen above, the sample size has generated just 18 vendors accounts covering 5% in terms of volume and 42% in terms of value.

6. Stratification

Stratification is a derivative of classification which involves grouping of large data into 'strata'. Strata means levels, bands or groups. Thus it involves dividing or rearranging the data within the Strata and then overviewing it to decipher the latent configuration of the database.

For example the purchase data of the above example can be stratified into various bands of bill values to obtain following results :

Strata		Bill Counts			Bill Value		
Bill_Val_From	Bill_Val_To	Count	Count%	Cumm%	Sum_Bill_Val	%	Cumm%
-	10,000	3081	93.85%	100.00%	34,49,671.74	30%	100%
10,001	20,000	77	2.35%	6.15%	11,14,929.05	10%	70%
20,001	30,000	45	1.37%	3.81%	11,35,250.64	10%	61%
30,001	40,000	37	1.13%	2.44%	13,47,267.34	12%	51%
40,001	50,000	13	0.40%	1.31%	6,01,560.32	5%	40%
50,001	1,00,000	14	0.43%	0.91%	9,17,721.94	8%	35%
1,00,001	1,50,000	7	0.21%	0.49%	8,68,979.09	7%	27%
1,50,001	3,00,000	7	0.21%	0.27%	14,40,783.65	12%	19%
3,00,001	4,00,000	1	0.03%	0.06%	3,19,000.00	3%	7%
4,00,001	5,00,000	1	0.03%	0.03%	4,88,352.19	4%	4%
		3283			1,16,83,515.95		

As can be seen, a sample size can be easily determined from above to mark bill values over Rs 50,000/- (depicted by dotted line in table above). This translates into a 0.91% in volume terms and 35% in value terms.

Strata can be quantities, insurance-policy holder-agewise, debtors ageing analysis, etc..

7. Single Transaction

This is self explanatory and may need little explanation. As the name suggests, this check basically filters all the single transactions in a database. These single records could be bonafide cases or just a stray transaction inserted by opportunist beneficiary. Generally vendor account, employee account, customer account, etc. should have multiple transactions since everyone wants regular business. Solitary transaction could be vouchsafe to exonerate sketchiness if any.

In a database of a medical distribution arm of a pharmaceutical company (Pharmaco), it was noticed that there were many stray records of doctors who were reimbursed train fares as part of promotional scheme. The promotional offer was to incentivize the doctors by allowing them tourism along with their family. The Pharmaco regularly had these schemes -- at lower end it was distribution of some token, gifts, etc. whereas at higher end senior doctors were allowed foreign travels, costly medical equipments, etc. The database analysis of doctor-wise expense of last 3-4years revealed these stray instances. While the individual vouchers did not throw up any irregularities, collectively all these single instances had a pattern of suspicious behavior. These vouchers were otherwise vouchedsafe with proper supportings (original train tickets), manager's approvals, names and particulars of the doctors, etc. However, collectively viewed the skepticism was obvious; train tickets of many beneficiaries pertained to journeys for same day with same coach numbers – this breached the test of impossibility. Field verification exposed fake names and addresses of many of these doctors, however the tickets appeared genuine. On detailed investigation, it was revealed that the accomplice would collect all the tickets from the Railway-ticket-checker by tipping him a good amount. Suitable tickets (family/groups) were later attached as supportings and monies siphoned off on regular basis. The fraud was perpetrated into small amounts on regular basis aggregating to large amount eventually. Clearly there were many single instances of genuine cases also but that actually facilitated in camouflaging the fraud.
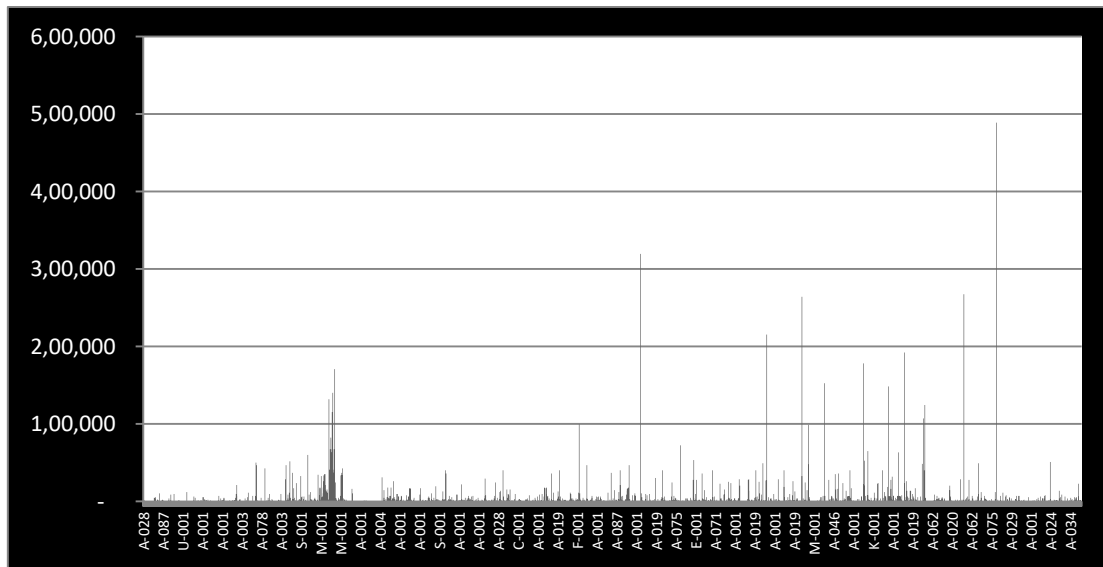
8. Isolated Outliers

An Isolated outlier is an observation in a data set which is far extreme in value from the others in the data set. It is an unusually large or an unusually small value compared to the others. Any database will be vitiated by incongruent records or contaminated transactions which will stick out as outliers. That happens because of its inherent nature that impedes its blending with the others in the group and will be clearly isolated with the remainders.

A word of caution -- there could outliers that would creep in any database as deviations which happens in normal course and may not always mean a fraud or an error. However, as an auditor s/he will be concerned about these outliers and should review these transactions as part of audit plan.

There are various ways to spot the Isolated Outliers as discussed below.

- Simple Charting Options : Using a popular spreadsheets, it is not difficult to plot a chart of the entire database as shown below. The database contains 3,283 records of 328 vendors. The vendor codes are plotted against the x-axis and the amount against the y-axis.

A simple visual review of the chart can enable the auditor to immediately spot the isolated outliers. The above chart is reproduced below with the dotted outline :
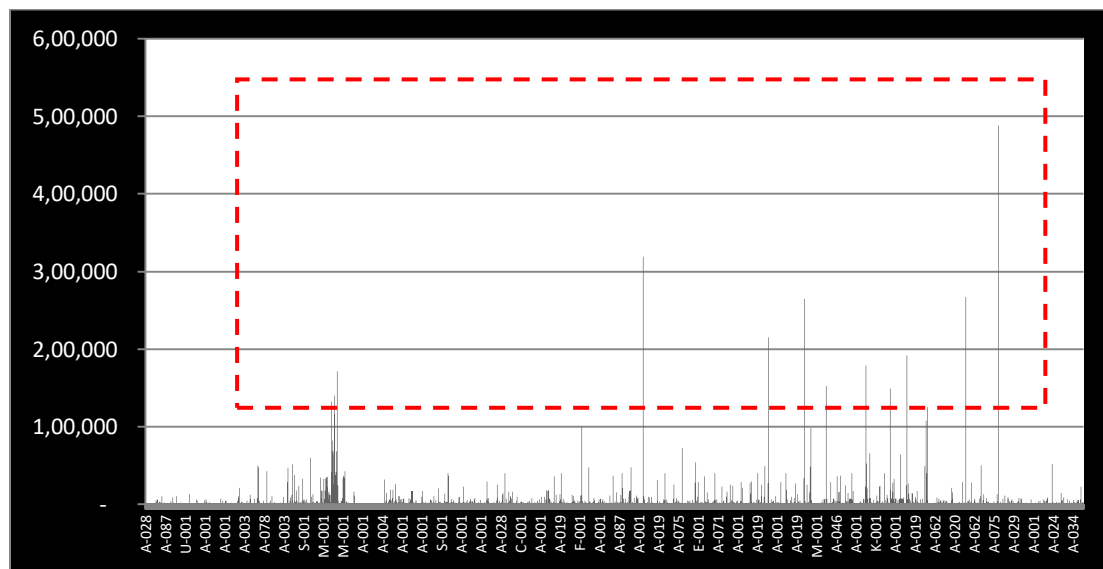
Thus all the transactions above Rs 1,00,000/- are isolated outliers.

- **Relative Size Factor :**

  RSF is the ratio of Largest Number to the Second Largest Number of a relevant set.

  $$RSF = \frac{Largest\ Number}{Second\ Largest\ Number}$$

  For example, if we have following bank payment vouchers of Vendor XYZ

Voucher No.	Rs.
SB-211	50,000
SB-642	5,00,000
SB-547	5,00,000
SB-1864	20,000
SB-4755	23,000
SB-8347	8,500

  The Largest value in above table = Rs 5,00,000/- and the second largest value = Rs 50,000/-. Therefore the RSF in this case = 10 that is Rs 5,00,000 Lacs divided by Rs. 50,000/-

  Per RSF theory generally any transactions where RSF > 10 are the cases of isolated outliers.

- **Benford's Law**

This is revolutionary theorem profounded by Dr. Frank Benford, as American Electrical Engineer and Physicists. Benford's Law is also popularly known as the first digit law. The law is about statistical statement regarding occurrence of numerical digits. Dr. Benford observed that in any large database generated through an ordinary process, the natural numbers (numbers which are not limited by boundaries or non-serial nos.) follow a count of its first-left-most digit which is not in consonance with the law of probability. He asserted that the first-leftmost-digit (e.g it is "1" in a number 1,23,456.78") follows a pattern of appearance where the lower numbers have more chance of appearing as compared to the higher numbers. According to him the appearance of first left most digit has the following frequency.

Digit	
1	30.10 %
2	17.60 %
3	12.50 %
4	9.70 %
5	7.90 %
6	6.70 %
7	5.80 %
8	5.10 %
9	4.60 %

Thus, numbers deviating from the above principle would be transactions that could are isolated outliers which could be easily identified using following example. Samples can be drawn from vouchers beginning with number 3 if one is reviewing say 1,000 purchase expenses invoices with following analysis using Benford's Law.

Digit	Bendford Law%	Expected Nos. (Blaw)	Actual Nos.	Variance	Deviation
A	B	C: B X 1000	D	E: D-C	F : E/C
1	30.10%	301	312	11	3.65%
2	17.60%	176	180	4	2.27%
3	12.50%	125	146	21	16.80%
4	9.70%	97	100	3	3.09%
5	7.90%	79	75	-4	-5.06%
6	6.70%	67	62	-5	-7.46%
7	5.80%	58	54	-4	-6.90%
8	5.10%	51	47	-4	-7.84%
9	4.60%	46	24	-22	-47.83%
	100.00%	1,000	1000	0	

For Free tutorials on MS Excel, please visit us on www.chetandalal.com and register for free eLearning course on how to use basic functions of MSExcel for audit and other general use. See FAQs on the foregoing website. This is limited period offer and subject to terms and conditions.

*****